

ABSTRACT

Systems and methods for analyzing boilerplate are described. In one described system, an indexer identifies a common element in a plurality of related articles. The indexer then classifies the common element as boilerplate. For example, the indexer may identify a copyright notice appearing in a plurality of related articles. The copyright notice in these articles is considered boilerplate.